Journal of Medical Virology 48:60–67 (1996)

Sequence and Genomic Organization of GBV-C: A Novel Member of the Flaviviridae Associated With Human Non-A-E Hepatitis

Thomas P. Leary, A. Scott Muerhoff, John N. Simons, Tami J. Pilot-Matias, James C. Erker, Michelle L. Chalmers, George G. Schlauder, George J. Dawson, Suresh M. Desai, and Isa K. Mushahwar

Virus Discovery Group, Experimental Biology Research, Abbott Laboratories, North Chicago, Illinois

Recently, sequences from a novel virus, termed GB virus C (GBV-C), were identified in serum from several patients with cryptogenic hepatitis. In the present study, the nucleotide sequence of this virus has been extended to near-genome length. GBV-C encodes a putative single large polyprotein in which the structural proteins are positioned at the N-terminal end, with the nonstructural proteins located at the C-terminal end. Amino acid sequence analysis of this large polyprotein reveals the presence of protease, helicase, and replicase motifs. Sequence alignments of the polyprotein followed by phylogenetic analyses suggest that GBV-C is a member of the Flaviviridae, most closely related to the recently described GB virus A.

KEY WORDS: GB virus, RNA virus, HCV

INTRODUCTION

Although reliable assays are available for distinguishing the non-A, non-B hepatitis viruses (hepatitis C and E viruses) [Kuo et al., 1989; Dawson et al., 1992; Schlauder and Mushahwar, 1994], approximately 10–20% of hepatitis cases are of unknown etiology. A number of clinical studies have suggested the existence of additional viral agents. These include individuals with elevated serum aminotransferase (ALT) levels [Marcellin et al., 1993], community-acquired acute resolving and chronic hepatitis [Alter et al., 1992; Tassopoulos et al., 1992; Buti et al, 1994], and fulminant hepatitis [Fagan et al., 1992; Kuwada et al., 1994; Sallie et al., 1994].

Among the candidate viruses for human hepatitis of unknown etiology are two novel viral genomes found in the GB hepatitis agent [Schlauder et al., 1995; Simons et al., 1995a]. These viruses, termed GB virus A (GBV-A) and GB virus B (G BV-B), have positive-strand, linear RNA genomes approximately 10 kB in length. Each possess single large open reading frames that appear to encode the structural proteins at the 5′ end of the genome, encoding the nonstructural proteins at the 3′ end. Based on amino acid sequence alignments followed by phylogenetic analyses, GBV-A and GBV-B appear to be members of the Flaviviridae [Muerhoff et al., 1995].

Degenerate PCR primers designed to amplify the helicase regions of GBV-A and GBV-B and a third member of the Flaviviridae, hepatitis C virus (HCV), were used to isolate the helicase gene sequence of a novel virus from the serum of several patients diagnosed with cryptogenic (non-A-E) hepatitis [Simons et al., 1995b]. Because of the high degree of identity with GBV-A (59% at the nucleotide level and 64% at the amino acid level), this virus was termed GB virus C (GBV-C). In the present paper, we report the sequence and genomic organization of this novel virus as well as the results of an analysis of the putative viral proteins. Phylogenetic analyses demonstrate that GBV-C is an additional member of the Flaviviridae, distinct from the HCV group.

MATERIALS AND METHODS

Genome Extension

The genome of the initial 322 bp GBV-C cDNA was extended sequentially as described previously [Simons et al., 1995a]. Briefly, serum total nucleic acids extracted from a West African patient were subjected to reverse transcription with random hexamers (GeneAmp RNA PCR Kit; Perkin-Elmer Corp., Norwalk, CT), then employed as template in a "nested" PCR reaction (Perkin-Elmer). The primary PCR (50 μl) employed a biotinylated gene-specific primer and a semi-random flanking primer [Sorensen et al., 1993]. Following 35 cycles of amplification (94°C, 30 sec; 55°C, 30 sec; 72°C, 90 sec; final extension at 72°C for

Accepted for publication July 18, 1995.

The sequence reported in this paper has been deposited in the GenBank database (Accession No. U36380).

Address reprint requests to Thomas P. Leary, Abbott Laboratories, Dept. 90D, Bldg. L3, 1401 Sheridan Road, North Chicago, IL 60064.

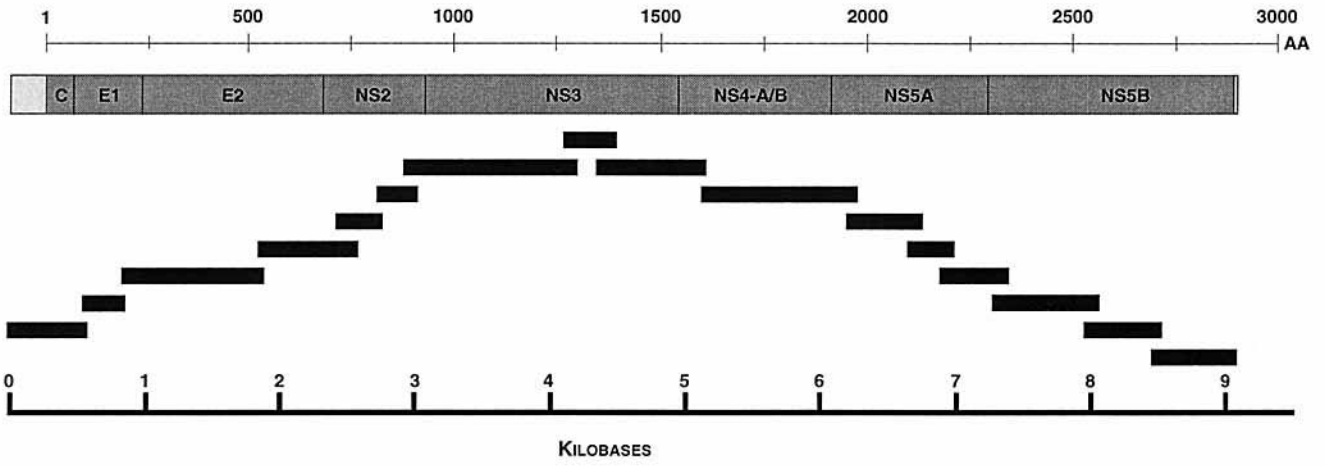

Fig. 1. Genomic extension clones and the genomic organization of GBV-C. Thick lines indicate the relative size and locations of the overlapping genomic extension clones. The dark shaded box designates the putative large open reading frame beginning with the proline residue at position one, while the lighter boxes indicate the presumed noncoding regions. Putative structural and nonstructural proteins are denoted.

10 min), biotinylated products were purified on streptavidin paramagnetic particles (Promega, Madison, WI), washed, and then eluted with NaOH [Sorensen et al., 1993]. The secondary PCR (50 μl) was carried out for 35 cycles utilizing 6.7% of the purified first-round product, a nested gene-specific primer and a nonrandom flanking primer [Sorensen et al., 1993]. Products of the secondary reaction were gel isolated, cloned into pT7Blue T-vector (Novagen, Madison, WI), and then subjected to double-strand DNA sequence analysis (Sequenase Version 2.0 Sequencing Kit; United States Biochemical, Cleveland, OH).

Computer Analysis

Nucleotide sequences were compiled using the Fragment Assembly Program of the Wisconsin Sequence Analysis Package (Version 8, September, 1994; Genetics Computer Group, Madison, Wisconsin). Sequence alignments were undertaken with the PILEUP program from the same package, and regional comparisons used the PLOTSIMILARITY program with the identity setting in place. Phylogenetic distances between pairs of the amino acid sequences were determined by the PROTDIST program from the PHYLIP package (version 3.5c) [Felsenstein, 1993]. These computed distances were used for the construction of phylogenetic trees using the program NEIGHBOR (neighbor-joining setting), with DRAWTREE producing the final output. Additional analyses utilized FITCH. GenBank accession numbers of viral sequences used in the alignments are as follows: GBV-A, U22303; GBV-B, U22304; HCV-1 (genotype 1a), M62321; HCVJK1 (genotype 1b), X61596; HCVJ6 (genotype 2a), D00944; HCVJ8 (genotype 2b), D10988; HCV3A (genotype 3a), D28917.

RESULTS

Extension of the GBV-C Genome

Previously, we demonstrated that the initial 322 bp cDNA fragment of GBV-C was exogenous to the genomes of human, *Saccharomyces cerevisiae,* and *Escherichia coli* and that it was present only in RNA forms [Simons et al., 1995b]. To establish that this cDNA fragment was part of a larger viral entity, sequences both up- and downstream of the helicase region were sequentially extended to 9,125 nucleotides in length (Fig. 1). The largest open reading frame deduced from the GBV-C sequence consists of 2,906 amino acids that begins with a proline residue (amino acid 1) and concludes with a translational termination codon. This is followed by 61 nucleotides of presumably untranslated sequence. Whereas a methionine residue is found 32 amino acids downstream from the proline residue, the initiator methionine has yet to be identified (see below). An additional 343 nucleotides reside upstream of the proline residue.

TABLE I. Amino Acid Sequence Identity (%) Amongst Flaviviridae Polyproteins*

	HCV-1	GBV-A	GBV-B
GBV-A	26		
GBV-B	32	27	
GBV-C	29	48	28

*Amino acid sequence alignments were carried out as described in Materials and Methods. Percentage identity is calculated by adding the number of identical pairs of amino acids in the alignment and dividing this figure by the length of the smaller of the two sequences where gaps have been omitted.

Comparisons of GBV-C to GBV-A, GBV-B, and HCV-1

Amino acid sequence comparisons of the deduced GBV-C large open reading frame to those of GBV-A, GBV-B, and HCV-1 reveal that GBV-C is most similar to GBV-A, being 48% identical (Table I). GBV-C is approximately 30% identical across this same region to

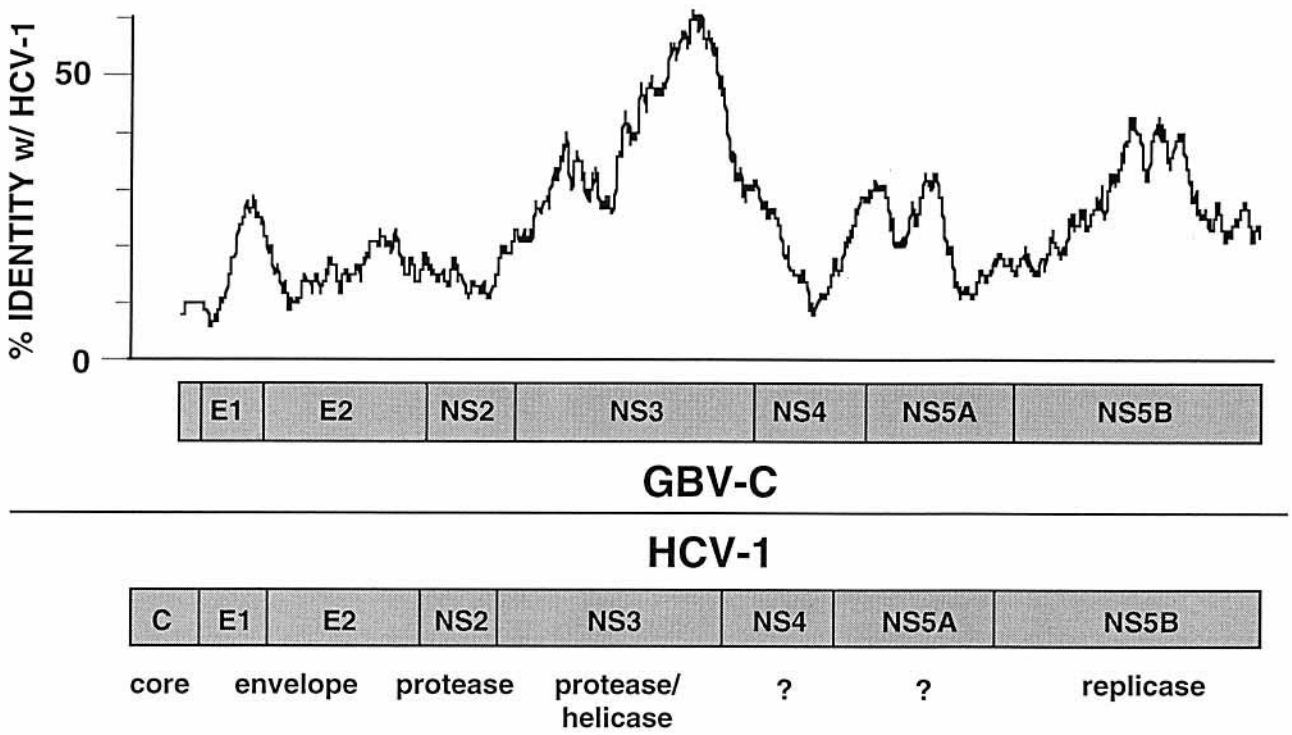

Fig. 2. Regional amino acid sequence identity of GBV-C compared to that of HCV-1. Following amino acid sequence alignment of the GBV-C and HCV-1 large open reading frames, the percentage identity of a 100-amino-acid sliding window of comparison was plotted. Shaded boxes represent the GBV-C and HCV-1 large open reading frames. The GBV-C open reading frame begins at position 1 with the proline residue and concludes with the translation termination codon. Alignments and plots are as described in Materials and Methods.

GBV-B and HCV-1. Based on this analysis, GBV-C is no more related to HCV-1 than is GBV-A or GBV-B (26% and 32%, respectively) and shares approximately the same identity with GBV-B as does GBV-A (28% vs. 27%). BLAST searches of the GenBank database with the GBV-C amino acid sequence reveal significant regional identity to HCV, in addition to limited regional identity with other positive-strand RNA viruses. These include bovine viral diarrhea virus, hog cholera virus, and several plant viruses.

A more detailed comparison of regional identity can be carried out utilizing an end-to-end amino acid alignment of two sequences, followed by plotting the percentage identity of small windows across the alignment. Such analysis on GBV-C vs. HCV-1 reveals several regions of local identity between these two sequences (Fig. 2). As was observed for Flavi-like viruses [Koonin and Dolja, 1993], the greatest region of identity occurs within the helicase region. Similar profiles are obtained when comparing GBV-C to either GBV-A or GBV-B (data not shown). Also suggestive of this region encoding a viral helicase, GBV-C possesses residues conserved in the supergroup II helicases of positive-strand RNA viruses (Fig. 3A) [Koonin and Dolja, 1993].

Lesser, but significant, amino acid sequence identity between GBV-C and HCV-1 occurs within the N-terminus of the putative NS3 region and in the putative NS5B region (Fig. 2). These regions of HCV encode a serine protease and an RNA-dependent RNA polymerase [Choo et al., 1991] and are also conserved in GBV-A and GBV-B [Muerhoff et al., 1995]. In HCV-1, the serine protease catalytic triad is composed of His, Asp, and Ser residues at positions 1083, 1107, and 1165, respectively. These residues are conserved and appropriately spaced in GBV-C (positions 1013, 1037, and 1094), suggesting that this novel virus also encodes a serine protease and in all likelihood processes its polyprotein in a manner similar to HCV. Within the putative NS5B region, a number of residues conserved in the supergroup II replicases of positive-strand RNA viruses [Koonin and Dolja, 1993] are maintained in GBV-C (Fig. 3B). These include the Gly-Asp-Asp (amino acids 2649–2651) signature sequence thought to be part of the catalytic domain of the replicase [Gorbalenya and Koonin, 1989]. Thus, it is probable that GBV-C also encodes an RNA-dependent RNA polymerase similar to other members of the Flaviviridae, including HCV, GBV-A, and GBV-B.

Additional, albeit less conserved, regions between GBV-C and HCV-1 are observed within the putative E1, NS4, and NS5A regions (Fig. 2). Similar regions of identity are found when comparing GBV-C to GBV-A or GBV-B (data not shown). Insofar as the functional role of the HCV NS4 and NS5A proteins is unknown, the significance of the observed conservation is unclear. However, based on the regional sequence identity between these viruses, it might be expected that these proteins perform a similar function.

A

```
GBV-C  RFLANPRQML RGVSVVICDE CHSHDSTVLL GIGRVRDVAR GCGVQLVLYA TATPPGSPMT QHPSIIETKL DV.GEIPFYG HGIPLERMRT GRHLVFCHSK
GBV-A  RFMANPRKYL RGNDVVICDE LHVTDPTSIL GMGRARLLAR ECGVRLLLFA TATPPVSPMA KHESIHEEML GSEGEVPFYC QFLPLSRYAT GRHLLFCHSK
GBV-B  MYL..TGACS RNYDVIICDE CHATDATTVL GIGKVLTEAP SKNVRLVVLA TATPPGVIPT PHANITEIQL TDEGTIPFHG KKIKEENLKK GRHLIFEATK
HCV-1  KFLADGGCSG GAYDIIICDE CHSTDATSIL GIGTVLDQAE TAGARLVVLA TATPPGSVTV PHPNIEEVAL STTGEIPFYG KAIPLEVIKG GRHLIFCHSK
Cons.  ---------- ------ICDE -H--D-T--L G-G-----A- -----L---A TATPP----- -H--I-E--L ---G--PF-- ---------- GRHL-F---K
                        **   *                                *****

GBV-C  AECERLAGQF SARGVNAIAY YRGKDSSIIK .DGDLVVCAT DALSTGYTGN FDSVTDCGLV VEEVVEVTLD PTITISLRTV PASAELSMQR RGRTGRGRSG
GBV-A  VECTRLSSAL ASFGVNTVVY FRGKETD.IP .TGDVCVCAT DALSTGYTGN FDTVTDCGLM VEEVVEVTLD PTITIGVKTV PAPAELRAQR RGRCGRGKAG
GBV-B  KHCDELANEL ARKGITAVSY YRGCDISKIP .EGDCVVVAT DALCTGYTGD FDSVYDCSLM VEGTCHVDLD PTFTMGVRVC GVSAIVKGQR RGRTGRGRAG
HCV-1  KKCDELAAKL VALGINAVAY YRGLDVSVIP TSGDVVVVAT DALMTGYTGD FDSVIDCNTC VTQTVDFSLD PTFTIETITL PQDAVSRTQR RGRTGRGKPG
Cons.  --C--L---- ---G-----Y -RG-----I- --GD--V-AT DAL-TGYTG- FD-V-DC--- V-------LD PT-T------ ---A----QR RGR-GRG--G
                                                * *   *                                                  **   ** **
```

B

```
GBV-C  LA...SDHPE WVRALG.KYY ASGTMVTPEG VPVGERYCRS SGVLTTSASN CLTCYIKVRA ACERIGLKNV SLLIAGDDCL IVCERPVCDP CEALGRTLAS
GBV-A  AA...SDNPS MVHALC.KYY SGGPMVSPDG VPLGYRQCRS SGVLTTSSAN SITCYIKVSA ACRRVGIKAP SFFIAGDDCL IIYENDGTDP CPALKAALAN
GBV-B  AAKLSDQHRA GIHTIARQYH AGGPMIAYDG REIGYRRCRS SGVYTTSSSN SLTCWLKVNA AAEQAGMKNP RFLICGDDCT VIWKSAGADA DKQAMRVFAS
HCV-1  CCDLDPQARV AIKSLTERLY VGGPLTNSRG ENCGYRRCRA SGVLTTSCGN TLTCYIKARA ACRAAGLQDC TMLVCGDDLV VICESAGVQE DAASLRAFTE
Cons.  ---------- ---------- --G------G ---G-R-CR- SGV-TTS--N --TC--K--A A----G---- -----GDD-- ---------- ----------
                                              #  ##   #   #                      ###
```

Fig. 3. Amino acid comparisons of helicase and replicase sequences. **A:** Alignment of the putative NS3 helicase domain of GBV-C (residues 1229–1426), GBV-A (residues 1242–1439), GBV-B (residues 1212–1408), and HCV-1 (residues 1298–1497). Residues conserved between the supergroup II RNA helicases [Koonin and Dolja, 1993] are designated by asterisks. **B:** Alignment of the putative replicases around the GDD signature sequence of GBV-C (residues 2578–2673), GBV-A (residues 2634–2729), GBV-B (residues 2513–2612), and HCV-1 (residues 2662–2761). Residues conserved in the supergroup II replicases [Koonin and Dolja, 1993] are designated by number signs. Alignments were carried out as described in Materials and Methods.

```
                                                                                         core⮢ ⮣E1               *
GBV-C  DQWGPGGRGKDPHRCPSRGGGKCMGPPSSAAAYSRGSPRTLRVRAGGISFPIPIMA......VLLLLLVVEAGAILAPA..THACSAKGQYFLTNCCALE
GBV-A  ......................PLVAIPS............LRSMSVVDTFT...MA.......WLWLLVCFPLAGGVLFNSRHQCFNGDHYVLSNCCSRD
GBV-B  RNLGILLDYPLGWIGDVTTHTPLVG.PLVAGAVVRPVCQIVRLLEDGVNWATGWF...GVHLFVVCLLSLACPCSGARVTDPDTNTT....ILTNCCQRN
HCV-1  RNLGKVIDTLTCGFADLMGYIPLVGAPLGGAA..RALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSTGL....YHVTNDCPNS
                                                                                       (signal sequence)

                  *  *         *                                                               E1⮢ ⮣E2
GBV-C  DIGFCLEGGCLVALGCTIC.....TDRCW........(120 aa).......PLLVCVAALLLLE.QRIVMVFLLVTMAGMSQGAPASVLGSRPFEA...
GBV-A  EVYFCFGDGCLVAYGCTVC.....TQSCW........(106 aa)........LSLLVILMLVLE.QRLLIAFLLLLVVGEAQRGMFDNCVCGYWGG...
GBV-B  QVIYCSPSTCLHEPGCVIC.....ADECW........(119 aa).......AVGALIYYASRGKWYQLLLALMLYIEATSGNPIRVPTGCSIAEFC...
HCV-1  SIVYEAADAILHTPGCVPCVREGNASRCW........(115 aa).......VLAGIAYFSMVGNWAKVLVVLLLFAGVDAETHVTGGSAGHTVSGF...
                                                                                 (signal sequence)
```

Fig. 4. Amino acid sequence alignment of core/E1 and E1/E2 regions at the putative eukaryotic signal sequence cleavage sites. The GBV-C sequence begins at amino acid 10 of the largest open reading frame, and the GBV-A sequence begins at amino acid number 1. HCV-1 and GBV-B sequences were truncated to facilitate alignment. Cysteine residues that are maintained among all four viruses are indicated by asterisks, and residues observed in all four viral sequences are shown in boldface. The putative signal peptidase recognition sites [Bukh et al., 1993] are underscored, with the cleavage sites occurring at the end of the underlined signal sequence.

Potential Proteolytic Processing of the GBV-C Polyprotein

In HCV, host and viral proteases are responsible for processing of the structural and nonstructural proteins, respectively. This is also presumed to be the case for GBV-A and GBV-B [Muerhoff et al., 1995]. Within the inferred structural region of GBV-C, four eucaryotic signal sequence cleavage sites [vonHeijne, 1986] are evident, spaced similarly to those of HCV [Hijikata et al., 1991; Lin et al., 1994]. This suggests that host proteases are also involved in GBV-C polyprotein cleavage analogous to that of HCV. The core/E1 cleavage site in GBV-C appears to be localized between the first eucaryotic signal sequence cleavage site, contained within the predicted core, and an Asn-Cys-Cys motif (amino acids 96–98) within the putative E1 that is highly conserved in GBV-A, GBV-B, GBV-C, and HCV-1 (Fig. 4). Within the inferred envelope region, three consensus eucaryotic signal sequence cleavage sites can be identified, suggesting that the envelope region is hydrolyzed by host signal protease into three distinct proteins; E1 (GBV-C amino acids 80–267), E2 (amino acids 268–

645), and p7 (644–704), as is the case with HCV [Lin et al., 1994; Mizushima et al., 1994].

Cleavage of the NS2/NS3 bond in HCV occurs by way of a viral encoded protease encompassing the C-terminal end of NS2 and the N-terminal end of NS3. His and Cys residues at positions 952 and 993 of HCV, respectively, have been shown to be essential for hydrolyzing the peptide bond between the NS2 and NS3 proteins [Grakoui et al., 1993a]. Though little amino acid sequence identity exists between GBV-C and HCV, GBV-A, or GBV-B in this region of the polyprotein, His and Cys residues are conserved in each of these viruses (positions 881 and 922 of GBV-C), suggesting a similar mode of processing.

Additional processing of the HCV nonstructural proteins (NS3–NS5B) occurs via a second NS3-encoded viral serine protease, where the scissile bond is flanked by Cys/Thr in the P1 position and Ser/Ala in the P1′ position. Additionally, an acidic residue is found at the P6 position [Grakoui et al., 1993b]. The presence of a serine protease motif encoded by GBV-C suggests that this virus also catalyzes the cleavage of downstream nonstructural proteins in a manner similar to HCV. Computer alignments of the GBV-C deduced amino acid sequence reveal several amino acid stretches that may serve as substrates for this putative enzyme (Fig. 5). At a region of GBV-C that aligns well with the NS3/NS4A junction of HCV-1, the predicted cleavage site is at a Thr-Asp (amino acids 1612 and 1613 of GBV-C) peptide bond. Though an NS4B/NS5A cleavage site cannot be determined by this analysis, potential NS4A/NS4B and NS5A/NS5B cleavage sites can be identified between amino acids 1823 and 1824 and between amino acids 2342 and 2343, respectively, of GBV-C (Fig. 5). Both of these regions are flanked by Ala and Ser at the predicted scissile bond, and good alignment is found between GBV-A, GBV-B and HCV-1. Unlike the case for HCV, in no instance was an acidic residue found in the P6 position of these putative cleavage sites. Experimental evidence with HCV suggests that the P6 residue is not required for efficient proteolysis [Bartenschlager et al., 1995]. Empirical data regarding proteolytic cleavage of the predicted peptide bonds is necessary for more definitive delineation of the GBV-C structural and nonstructural proteins.

NS3/NS4A:

6	1 1′	
CMSADLEVV	**T S**	TWVLVGGVL HCV-1
HWTSSLVVV	**T S**	WVVNGNGNP GBV-A
YQMCFTEVN	**T S**	GTAALAVGV GBV-B
SYTGSLVVV	**T D**	WDVKGGGNP GBV-C

NS4A/NS4B

YFEFDEMEE	**C S**	QHLPYIEQG HCV-1
GGWEGAVNA	**A S**	LTFDLLAGK GBV-A
VDEEEIVEE	**C A**	SFIPLEAMV GBV-B
GGWEGVVNA	**A S**	LVFDFMAGK GBV-C

NS5A/NS5B

EANAEDVVC	**C S**	MSYSWTGAL HCV-1
CDQIEEETP	**T S**	YSYIWSGAP GBV-A
LGKSE--FS	**C S**	MSYTWTDVI GBV-B
EARQETL--A	**S**	FSYIWSGVP GBV-C

Fig. 5. Predicted cleavage sites for the putative GBV-C serine protease. The predicted NS3/NS4A, NS4A/NS4B, and NS5A/NS5B cleavage sites in the GBV-C polyprotein are displayed. The P6 (6), P1 (1), and P1′ (1′) positions around the scissile bond are designated, and the amino acid residues flanking the bond are shown in boldface.

Phylogenetic Analysis of GBV-C

As stated previously, limited amino acid sequence identity exists between the three GB viruses and HCV-1 (Table I). To determine more accurately the degree of relatedness between these viruses, amino acid sequence alignments were performed utilizing the predicted large open reading frames of GBV-A, -B, and -C and the large open reading frames of several representative HCV isolates. From these alignments, phylogenetic distances were calculated to determine the evolutionary relationship of these viruses (Table II). Judging by this analysis, GBV-C is most closely related to GBV-A (0.928 amino acid replacements per position) and most distantly related to GBV-B (2.47 replacements per position). The minimal distance between GBV-C and any of the HCV genotypes is 2.116, and the maximal distance between any of the HCV genotypes is 0.396. Thus, as was the case for GBV-A and GBV-B [Muerhoff et al., 1995], the GBV-C sequence does not fall within the diversity of the HCV genotypes and, therefore, cannot be considered a genotype of HCV. In fact, the diversity observed between GBV-C and the closest HCV genotype is 5.3 times the maximal distance between any of the HCV genotypes. Additionally, the high degree of divergence between GBV-C and GBV-B indicates that these viruses are not genotypes of one another. Although GBV-A and GBV-C appear to be related, the data do not support the hypothesis that they are genotypes of the same virus.

The evolutionary distances shown in Table II were used to generate an unrooted phylogenetic tree (Fig. 6). The HCV genotypes are tightly grouped on a major branch of the tree, whereas GBV-B stands alone on a second major branch. GBV-C and GBV-A are present on a third major branch, though these viruses further

TABLE II. Evolutionary Distances of Flaviviridae Polyproteins*

	GBV-A	GBV-C	HCV-J6	HCV-J8	HCV-1	HCV-JK1	HCV-3A
GBV-C	0.928						
HCV-J6	2.412	2.149					
HCV-J8	2.412	2.163	0.179				
HCV-1	2.388	2.116	0.359	0.365			
HCV-JK1	2.408	2.127	0.366	0.379	0.174		
HCV-3A	2.443	2.158	0.388	0.396	0.325	0.323	
GBV-B	2.688	2.470	1.699	1.687	1.712	1.717	1.738

*Evolutionary distances were calculated using the PRODIST program as described in Materials and Methods.

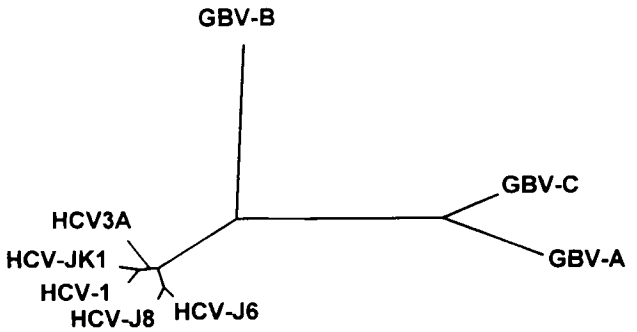

Fig. 6. Phylogenetic analysis of Flaviviridae polyproteins. The evolutionary distances in Table II were used to generate the unrooted phylogenetic tree as described in Materials and Methods.

diverge from a common ancestor. Interestingly, GBV-A is almost twice as far diverged from the common ancestor as is GBV-C. It is expected that, as additional isolates of each virus are found, they will form tightly clustered groups as is the case with the HCV genotypes.

DISCUSSION

In the present paper we describe the genomic sequence and organization of a novel virus present in the serum of a number of individuals diagnosed with cryptogenic hepatitis. This virus possesses a genome in excess of 9,100 nucleotides and appears to encode a single large polyprotein. This polyprotein encodes amino acid sequence motifs consistent with a serine protease, a helicase, and an RNA-dependent RNA polymerase. Finally, phylogenetic analysis of the amino acid sequence alignments demonstrate that GBV-C is a member of the Flaviviridae.

The genome of GBV-C appears to be organized much like those of HCV, GBV-A, and GBV-B. Comparison of the polyprotein reveals sequences consistent with structural proteins at the N-terminus, while the relative locations of the nonstructural protease, helicase, and replicase proteins are analogous to other members of the Flaviviridae. Areas of conservation in regions important for proteolytic processing further establish a common genomic organization between these viruses. Based on this observation, it is assumed that the GBV-C genome would be of positive polarity, although we have yet to demonstrate this experimentally due to limiting volumes of serum/plasma and the lack of biopsy material from patients infected with GBV-C.

Nucleic and amino acid sequences comparisons of GBV-C with those of HCV-1, GBV-A, and GBV-B indicate that the 5′ structural gene sequences are most similar to those of GBV-A. The nucleocapsid/E1 junction is predicted to occur approximately 46 amino acids downstream of the first methionine residue in the large open reading frame, and the open reading frame extends only an additional 32 amino acids upstream of this methionine residue. In comparison to other Flaviviridae nucleocapsid proteins, the open reading frame for the GBV-C nucleocapsid appears quite small. Both HCV and GBV-B have profoundly basic nucleocapsid proteins in excess of 150 amino acids, with calculated pI values of 11.9 and 11.1, respectively. The calculated pI of the 79 amino acids directly upstream of the predicted GBV-C nucleocapsid/E1 cleavage site is 11.77, suggesting that these sequences are part of the virus nucleocapsid. Sequences just downstream of the predicted cleavage site assume a near neutral pI, as is the case in both HCV and GBV-B. In excess of 300 nucleotides are present upstream of the large open reading frame. It is unclear whether these sequences further encode the nucleocapsid and the true reading frame is disrupted by sequence compressions due to the high GC content, or whether these sequences represent the 5′ noncoding region, with the remainder of the nucleocapsid sequences being "looped out" during cDNA synthesis and/or PCR amplification. Of note is the fact that these nucleotides do not contain sequences consistent with HCV or GBV-B 5′ noncoding regions [Han et al., 1991; Muerhoff et al., 1995]. Finally, it is possible that this sequence represents the true status of GBV-C in this particular sample. Studies to verify and possibly extend the 5′-most sequences of GBV-C are underway.

The large open reading frame of GBV-C concludes with a translation termination codon, followed by an additional 61 nucleotides of untranslated sequence. 3′-rapid amplification of cDNA ends (RACE) experiments have failed to identify either poly-(A) or poly-(U) tracts in this region. In this regard, GBV-C closely resembles GBV-A [Muerhoff et al., 1995]. Most simple, positive-strand RNA viruses conclude with a poly-(A) tract, whereas HCV concludes with either a poly-(A) [Choo et al., 1991; Han et al., 1991] or poly-(U) [Okamoto et al., 1991] tract. GBV-B is unique in the it contains an additional 50 nucleotides downstream of a poly-(U) tract [Muerhoff et al., 1995]. Studies are in progress to define further the noncoding sequences of GBV-C.

Phylogenetic analyses address the divergence of the GBV sequences from the HCV sequences, comparing this result with the observed divergence among the HCV sequences. In particular, are the GBV sequences no further diverged from the HCV sequences than the HCV sequences are from one another? This condition is not met in the present data. As is shown in Table II, the minimal HCV-GBV distance (2.116 replacements per position) is more than five times the maximal HCV-HCV distance (0.396). Thus, the data do not support the hypothesis that the GBV sequences are members of a group whose diversity is delimited by members of the HCV group. Assuming that the HCV sequences utilized in this study are representative of the most divergent of the HCV genotypes, these results indicate that GBV-A, -B, and -C are not genotypes of HCV. Phylogenetic analysis of aligned regions within the viral replicase from representatives of HCV genotypes 1–5 with the homologous regions of GBV-A, -B, and -C confirm this result (data not shown). Furthermore, it appears that GBV-A and GBV-C are more closely related to each other than either is to GBV-B (Fig. 6), which suggests that GBV-A and GBV-C may be representatives of a separate viral lineage. This hypothesis is also supported by phylogenetic analysis of the helicase and replicase regions of these viruses (data not shown). To evaluate further the relationship of GBV-A and GBV-C, further isolates of each virus will be necessary for study of the genetic diversity within each virus group.

REFERENCES

Alter MJ, Margolis HS, Krawczynski K, Judson FN, Mares A, Alexander WJ, Hu PY, Miller JK, Gerber MA, Sampliner RE, Meeks EL, Beach MJ (1992): The natural history of community acquired hepatitis C in the United States. New England Journal of Medicine 327:1899–1905.

Bartenschlager R, Ahlborn-Laake L, Yasargil K, Mous J, Jacobsen H (1995): Substrate determinants for cleavage in cis and in trans by the hepatitis C virus NS3 proteinase. Journal Virology 69:198–205.

Bukh J, Purcell RH, Miller RH (1993): At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proceedings of the National Academy of Sciences USA 90:8234–8238.

Buti M, Jardi R, Rodriguez-Frias F, Quer J, Esteban R, Guardia J (1994): Etiology of acute sporadic hepatitis in Spain: The role of hepatitis C and E viruses. Journal of Hepatology 20:589–592.

Choo Q-L, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby A, Barr PJ, Weiner AJ, Bradley DW, Houghton M (1991): Genetic organization and diversity of the hepatitis C virus. Proceedings of the National Academy of Sciences USA 88:2451–2455.

Dawson GJ, Chau KH, Cabal CM, Yarbough PO, Reyes GR, Mushahwar IK (1992): Solid-phase enzyme-linked immunosorbent assay for hepatitis E virus IgG and IgM antibodies utilizing recombinant antigens and synthetic peptides. Journal of Virological Methods 38:175–186.

Fagan EA, David SE, Tovey GM, Lloyd G, Smith HM, Portman B, Tan K-C, Zuckerman AJ, Williams R (1992): Toga Virus-like particles in acute liver failure attributed to sporadic non-A, non-B hepatitis and recurrence after liver transplantation. Journal of Medical Virology 38:71–77.

Felsenstein J (1993): PHYLIP Inference Package, Version 3.5c. Dept. of Genetics, University of Washington, Seattle, Distributed by the author.

Gorbalenya AE, Koonin EV (1989): Virus proteins containing the purine NTP-binding pattern. Nucleic Acids Research 17:8413–8440.

Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM (1993a): A second hepatitis C virus-encoded proteinase. Proceedings of the National Academy of Sciences USA 90:10583–10587.

Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM (1993b): Characterization of the hepatitis C virus-encoded serine proteinase: Determination of proteinase-dependent polyprotein cleavage sites. Journal of Virology 67:2832–2843.

Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea MS, Tekamp-Olson P, Kuo G, Choo Q-L, Houghton M (1991): Characterization of the terminal regions of hepatitis C viral RNA: Identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end. Proceedings of the National Academy of Sciences USA 88:1711–1715.

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K (1991): Gene mapping of the putative structural region of the hepatitis C virus genome by in vitri processing analysis. Proceedings of the National Academy of Sciences USA 88:5547–5551.

Koonin EV, Dolja VV (1993): Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. Critical Reviews in Biochemistry and Molecular Biology 28:375–430.

Kuo G, Choo Q-L, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, Tegtmeier GE, Bonino F, Colombo M, Lee W-S, Kuo C, Berger K, Shuster JR, Overby LR, Bradley DW, Houghton M (1989): An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science 244:362–364.

Kuwada SK, Patel VM, Hollinger B, Lin HJ, Yarbough PO, Wiesner RH, Kaese D, Rakela J (1994): Non-A, non-B fulminant hepatitis is also non-E and non-C. American Journal of Gastroenterology 89: 57–61.

Lin C, Lindenbach BD, Pragai BM, McCourt DW, Rice CM (1994): Processing in the hepatitis C virus E2-NS2 region: Identification of p7 and two distinct E2-specific products with different C termini. Journal of Virology 68:5063–5073.

Marcellin P, Martinot-Peignoux M, Gabriel F, Branger M, Degott C, Elias A, Xu LZ, Larzul D, Erlinger S, Benhamou JP (1993): Chronic non-B, non-C hepatitis among blood donors assessed with HCV third generation tests and polymerase chain reaction. Journal of Hepatology 19:167–170.

Mizushima H, Hijikata M, Asabe S, Hirota M, Kimura K, Shimotohno K (1994): Two hepatitis C virus glycoprotein E2 products with different C termini. Journal of Virology 68:6215–6222.

Muerhoff AS, Leary TP, Simons JN, Pilot-Matias TJ, Dawson GJ, Erker JC, Chalmers ML, Schlauder GG, Desai SM, Mushahwar IK (1995): Genomic organization of GBV-A and GBV-B: Two new members of the Flaviviridae associated with GB-agent hepatitis. Journal of Virology 69:5621–5630.

Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka, Machida A, Miyakawa, Mayumi M (1991): Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: Comparison with reported isolates for conserved and divergent regions. Journal of General Virology 72:2697–2704.

Sallie R, Silva AE, Purdy M, Smith H, McCaustland K, Tibbs C, Portmann B, Eddleston A, Bradley D, Williams R (1994): Hepatitis C and E in non-A non-B fulminant hepatic failure: A polymerase chain reaction and serological study. Journal of Hepatology 20: 580–588.

Schlauder GG, Mushahwar IK (1994): Detection of hepatitis C and E virus by the polymerase chain reaction. Journal of Virological Methods 47:243–253.

Schlauder GG, Dawson GJ, Simons JN, Pilot-Matias TJ, Guiterrez RA, Heynen CA, Knigge MF, Kurpiewski GS, Leary TP, Muerhoff AS, Desai SM, Mushahwar IK (1995): Molecular and serologic analysis in the transmission of the GB hepatitis agents. Journal of Medical Virology 46:81–90.

Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, Muerhoff AS, Erker JC, Buijk SL, Chalmers ML, Van Sant CL, Mushahwar IK (1995a): Identification of two flavivirus-like genomes in the GB hepatitis agent. Proceedings of the National Academy of Sciences USA 192:3401–3405.

Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, Schlauder GG, Desai SM, Mushahwar IK (1995b): Isolation of novel virus-like sequences associated with human hepatitis. Nature Medicine 1:564–569.

Sorensen AB, Duch M, Jorgensen P, Pedersen FS (1993): Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. Journal of Virology 67:7118–7124.

Tassopoulos NC, Hatzakis A, Delladetsima I, Koutelou MG, Todoulos A, Miriagou V (1992): Role of hepatitis C virus in acute non-A, non-B hepatitis in Greece: A five-year prospective study. Gastroenterology 102:969–972.

vonHeijne G (1986): A new method for predicting signal sequence cleavage sites. Nucleic Acids Research 14:4683–4690.